THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 261, No. 4, Issue of February 5, pp. 1778-1781, 1986
Printed in U.S.A.

Nucleotide Sequence of the *Escherichia coli dnaJ* Gene and Purification of the Gene Product*

(Received for publication, May 28, 1985)

Misao Ohki, Fumie Tamura, Susumu Nishimura, and Hisao Uchida‡

From the Biology Division, National Cancer Center Research Institute, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104 and the ‡Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

The *dnaJ* and *dnaK* genes are essential for replication of *Escherichia coli* DNA, and they constitute an operon, *dnaJ* being downstream from *dnaK*. The amount of the *dnaJ* protein in *E. coli* is substantially less than that of the *dnaK* protein, which is produced abundantly. In order to construct a system that overproduces the *dnaJ* protein, we started our study by determining the DNA sequence of the entire *dnaJ* gene, and an operon fusion was constructed by inserting the gene downstream of the λP_L promoter of an expression vector plasmid, pPL-λ. Cells containing the recombinant plasmid produced *dnaJ* protein amounting to 2% of the total cellular protein when cells were induced. The overproduced protein was purified, and Edman degradation of the protein indicated that the NH_2-terminal methionine was found to be processed. From the DNA sequence of the *dnaJ* gene, the processed gene product is composed of 375 amino acid residues, and its molecular weight is calculated to be 40,975.

DNA replication of bacteriophage λ requires functions of the phage-encoded proteins O and P, as well as those supplied by the host cell, *Escherichia coli* K12 (1, 2). The latter functions include those of the *dnaJ* and *dnaK* gene products (3–5), both of which are also essential for cell growth and appear to be related to cellular DNA and RNA synthesis (3, 6, 7). The *dnaJ* and *dnaK* genes constitute an operon (6) and are located in a region between *thr* and *leu* on the *E. coli* chromosome (5).

The properties of the *dnaK* gene product have recently attracted the interest of many investigators. The *dnaK* protein is identified as one of the heat shock proteins (8), and its DNA sequence is conserved among a wide variety of organisms ranging from procaryotes to humans (9). The *dnaK* protein possesses both ATPase and autophosphorylating activities (10). It is demonstrated from the *in vitro* studies cited above that the *dnaK* gene product directly interacts with the P protein.

On the other hand, information on the *dnaJ* gene product is very limited. Although the *dnaK* gene product is produced abundantly in *E. coli* and the *dnaJ* gene is located downstream of the *dnaK* gene constituting an operon, the amount of the *dnaJ* protein is substantially less than that of the *dnaK* protein. Therefore, we started our study by determining the DNA sequence of the *dnaJ* gene and constructing a recombinant plasmid that overproduces the *dnaJ* gene product.

EXPERIMENTAL PROCEDURES

Bacterial and Phage Strains and Plasmids—Recombinant phages λ*dnaKdnaJ* and λ*dnaKdnaJ*Δ27 were described elsewhere (6). Bacteriophages M13mp10 and M13mp11 and their host strain JM101 were kindly provided by Y. Kuchino (National Cancer Center Research Institute, Tokyo). A plasmid pMCR561 was kindly donated by T. Miki (Yamaguchi University, School of Medicine, Japan) (11). An expression plasmid pPL-λ that carries the P_L promoter and N gene on a 1215-base pair (bp[1]) segment of the genome inserted between the *Eco*RI and *Bam*HI site of pBR322 and its host strain N4830 (12) were obtained from Pharmacia/P-L Biochemicals.

Enzymes and Reagents—Various DNA-modifying and restriction enzymes were commercial products. [α-^{32}P]dATP (>400 Ci/mmol, 1 Ci = 37 GBq) was purchased from Amersham Corp. Dideoxy-NTPs and deoxy-NTPs were obtained from P-L Biochemicals and Sigma, respectively. Other reagents were commercial products of analytical grade.

Estimation of the dnaJ Protein—Samples were electrophoresed on a 2% $NaDodSO_4$, 12.5% acrylamide, 0.3% *N,N'*-methylenebisacrylamide slab gel, and the amount of the *dnaJ* protein was estimated by densitometry of protein bands stained with Coomassie Brilliant Blue. The *dnaJ* protein band was identified by superimposing on the gel the autoradiogram of the co-migrating *dnaJ* protein extracted from UV-irradiated cells infected with λ*dnaJ* phages incubated in the presence of [^{14}C]leucine. As noted by Georgopoulos *et al.* (13), the *dnaJ* protein behaves anomalously as migration is greatly affected by the acrylamide concentration of the gel. Under the condition employed in this study, the *dnaJ* protein migrated as a 40-kDa protein.

Construction of a Plasmid pDNAJ-A and an Expression Plasmid pPL-dnaJ-23—Thirty micrograms of λ*dnaKdnaJ*Δ27 DNA was digested with *Ava*I. By comparison of the digest with similar digests of the wild-type λ DNA and from knowledge on the structure of λ*dnaKdnaJ*Δ27, a 3.33-kilobase pair (kb) fragment was identified as containing the intact *dnaJ* gene. It was purified by agarose gel electrophoresis and ligated with molar excess of synthetic *Eco*RI linkers as described previously (14). The resulting fragments were inserted at the *Eco*RI site of plasmid pMCR561. Plasmid pDNAJ-A is one of the constructs selected randomly among ampicillin-resistant transformants. Thirty-five micrograms of pDNAJ-A was linearlized with *Sal*I and treated with 1.75 units of *Bal*31 exonuclease at 30 °C for 1.0 min. The mixture was then treated with *Eco*RI and fractionated by agarose gel electrophoresis, and the 2.65-kb band was extracted. The ends of the DNA fragments were filled in by treating with DNA polymerase I large fragment, and the resulting DNA was inserted into the *Hpa*I site of the pPL-λ DNA. The mixture of recombinant plasmids thus obtained was used to transform strain N4830 at 30 °C and ampicillin-resistant colonies were selected. Each transformant was cultured at 30 °C in broth. At a bacterial concentration of about 2×10^8/ml, the culture was transferred to a 42 °C bath. After incubation for 3.5 h, cell extracts were prepared and analyzed by $NaDodSO_4$-polyacrylamide gel electrophoresis. Nine out of 36 transformants examined produced various amounts of the *dnaJ* protein ranging from 0.3 to 2% of total cellular proteins as estimated

* This work was supported in part by grants from the Ministry of Education, Science, and Culture and the Special Coordination Fund from the Science Technology Agency of the Japanese Government. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] The abbreviations used are: bp, base pair; kb, kilobase pair; $NaDodSO_4$, sodium dodecyl sulfate; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

by densitometry of the stained gel. The clone which produced the largest amount of the *dnaJ* protein was named pPL-*dnaJ*-23.

Purification of the dnaJ Protein—Cells of strain N4830 harboring pPL-*dnaJ*-23 were grown at 30 °C in 2 liters of LB broth containing ampicillin (50 μg/ml). At a cell density of $OD_{560\,nm} = 0.4$, the culture was transferred to 42 °C and incubated for 4 h. The cells were harvested, suspended in a buffer containing 25 mM Hepes (pH 7.6), 1 mM EDTA, and stored at −70 °C. Cells were lysed as described (15); frozen-cell suspensions were thawed at 4 °C, adjusted to 80 mM KCl, 2 mM dithiothreitol, and 0.3 mg/ml egg white lysozyme was added. The cells were again frozen and thawed, and lysates were sonicated (30 s) three to four times. After removing cell debris by low speed centrifugation (3,000 × *g*, 10 min), the lysates were subjected to a high speed centrifugation (200,000 × *g*, 30 min). The sedimented proteins were suspended in a small volume of 50 mM potassium phosphate at pH 7.0, and NaCl was added to a concentration of 1 M. The mixture was left standing at 0 °C for 60 min, and insoluble materials were removed by centrifugation (200,000 × *g*, 60 min). The supernatant, which contained a large part of the *dnaJ* protein, was dialyzed against 50 mM potassium phosphate (pH 7.0), 6 mM β-mercaptoethanol. The precipitate which appeared during dialysis was collected by centrifugation (200,000 × *g*, 60 min) and suspended in the same buffer. Aliquots of this suspension were subjected to Na-DodSO_4-polyacrylamide gel electrophoresis. The *dnaJ* protein was eluted electrophoretically from slices as described (16).

Other Methods—DNA sequencing was done by the dideoxychain termination method of Sanger *et al.* (17). Restriction fragments were cloned into the M13mp10 or M13mp11 vectors for dideoxy sequencing (18). The procedures for preparing samples for Na-DodSO_4-polyacrylamide slab gel electrophoresis were as described by Ames (19). Purified proteins were sequenced by Edman degradation using an automated gas-phase sequencer (Applied Biosystems, Model 470A).

RESULTS

Nucleotide Sequence of the dnaJ Gene—Heteroduplex and complementation analyses indicated that the λ*dnaKdnaJ*Δ27 phage retains the *dnaJ* gene but lacks a 4-kb DNA fragment covering a large part of the *dnaK* gene and some of the bacterial DNA present in λ*dnaKdnaJ* (6). *Ava*I digestion of the λ*dnaKdnaJ*Δ27 produced a 3.3-kb fragment that is not present in the wild-type λ phage. The fragment was identified as containing the *dnaJ* gene (Fig. 1) and was cloned into pMCR561 to obtain pDNAJ-A (see "Experimental Procedures"). Digestion of the 3.3-kb fragment with *Pst*I produced 2.4-, 0.45-, and 0.44-kb DNA fragments. Both the 0.45- and 0.44-kb fragments are present in the wild-type λ DNA. The 2.4-kb fragment contains three *Hin*dIII sites, one of which is the site in the original vector DNA used to clone the *dnaK* gene (20) and is outside the Δ27 deletion (6). Therefore, the nucleotide sequence of the remaining region bracketed by the *Hin*dIII and a *Pst*I site was determined (Fig. 1). The sequencing revealed that the bacterial DNA fragment is present in λ*dnaKdnaJ*Δ27 DNA inserted between positions 20,651 and 27,479 (the *Hin*dIII site) of the wild-type λ DNA (21).

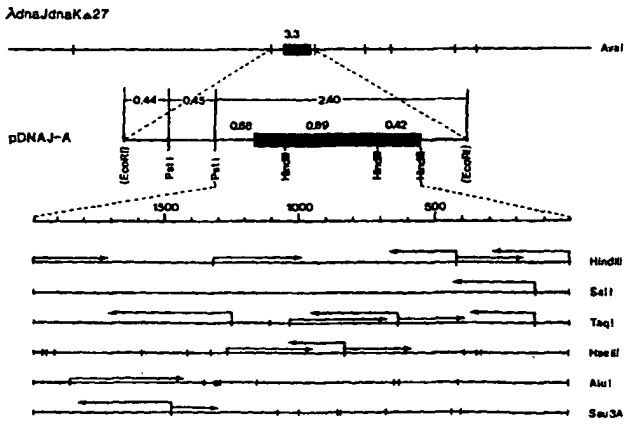

FIG. 1. **Strategies employed for DNA sequence analyses of the *dnaJ* gene of *E. coli*.** *Solid bars* in λ*dnaKdnaJ*Δ27 and pDNAJ-A represent the cellular DNA. A 1.9-kb *Pst*I-*Hin*dIII fragment was subcloned into M13mp10 or M13mp11 using restriction enzymes as indicated. *Arrows* show direction and extent of each DNA sequencing. About 75% of the final sequence was determined from both strands.

Fig. 2 shows the complete nucleotide sequence of the 1623-bp long region containing the *dnaJ* gene. A comparison of our sequence and the sequence of the *dnaK* gene as published by Bardwell and Craig (9) revealed that the sequence in the region of nucleotides 29–175 is identical to that for the COOH-terminal 49 amino acids of the *dnaK* protein. The site marked by an arrow in Fig. 2 is the site of the Δ27 deletion. An open reading frame is found in the region between nucleotides 267 and 1394, providing a coding capacity of 376 amino acids corresponding to a protein having a molecular weight of 41,106. Fig. 2 shows the deduced amino acid sequence.

Purification of the Overproduced dnaJ Protein— The *dnaK* protein is one of the most abundant protein species produced by *E. coli* and is easily detected by Na-DodSO_4-polyacrylamide gel electrophoretic analysis of total cellular proteins (8). Furthermore, the *dnaK* protein is known to be a heat shock protein that is produced even more abundantly when the bacteria are incubated at 42 °C (8). Although we have previously presented genetic evidence that the *dnaK* and *dnaJ* genes constitute an operon, the *dnaK* gene being located upstream of the *dnaJ* gene, the *dnaJ* protein is difficult to identify on the stained gel even after heat shock treatment.

A close examination of the DNA sequence of the *dnaJ* gene (Fig. 2) revealed a putative palindromic structure between the *Sal*I site and the initiation codon of the open reading frame. Therefore, in order to construct a plasmid that overproduces the *dnaJ* protein, the pDNAJ-A DNA was linearized by digestion with *Sal*I, digested by *Bal*31 exonuclease, and inserted into the expression vector, pPL-λ, at the *Hpa*I site (see "Experimental Procedures"). Strain N4830 containing one such recombinant plasmid, named pPL-*dnaJ*-23, produced *dnaJ* protein amounting to 2% of total cellular protein upon heat induction by incubation at 42 °C for 4 h (Fig. 3). When pPL-*dnaJ*-23 DNA was digested with *Hae*III, a 320-bp DNA fragment was detected instead of the 402-bp fragment which would be produced from pPL-λ DNA by the same treatment. Thus, the 82-bp region starting from the *Sal*I site and ending at the center of the putative palindromic structure was removed by the *Bal*31 digestion. The strain N4830 containing pPL-*dnaJ*-23 was used to purify the *dnaJ* protein (see "Experimental Procedures").

When the cell extracts were centrifuged and separated into soluble and insoluble fractions, most of the *dnaJ* protein was recovered in the insoluble fraction. The protein was eluted from the pellet at high concentrations of salts such as NaCl, KCl, or NH_4Cl; the addition of NaCl at a concentration of 1 M solubilized more than 80% of the *dnaJ* protein present in the precipitate. Dialysis of the eluate against a low concentration of NaCl (0.05 M) resulted in the appearance of a precipitate that is enriched for the *dnaJ* protein. Further purification of the protein was achieved by Na-DodSO_4-polyacrylamide gel electrophoresis and electrophoretic elution of the protein from the gel slices. This final preparation was nearly homogeneous, as shown in Fig. 3, *lane 9*.

Automated Edman degradation of the purified *dnaJ* protein revealed Ala-Lys-Gln-Asp-Tyr- as the amino-terminal sequence of the protein. The amino-terminal methionine was not detected, presumably due to *in situ* processing which is known to occur in many proteins.

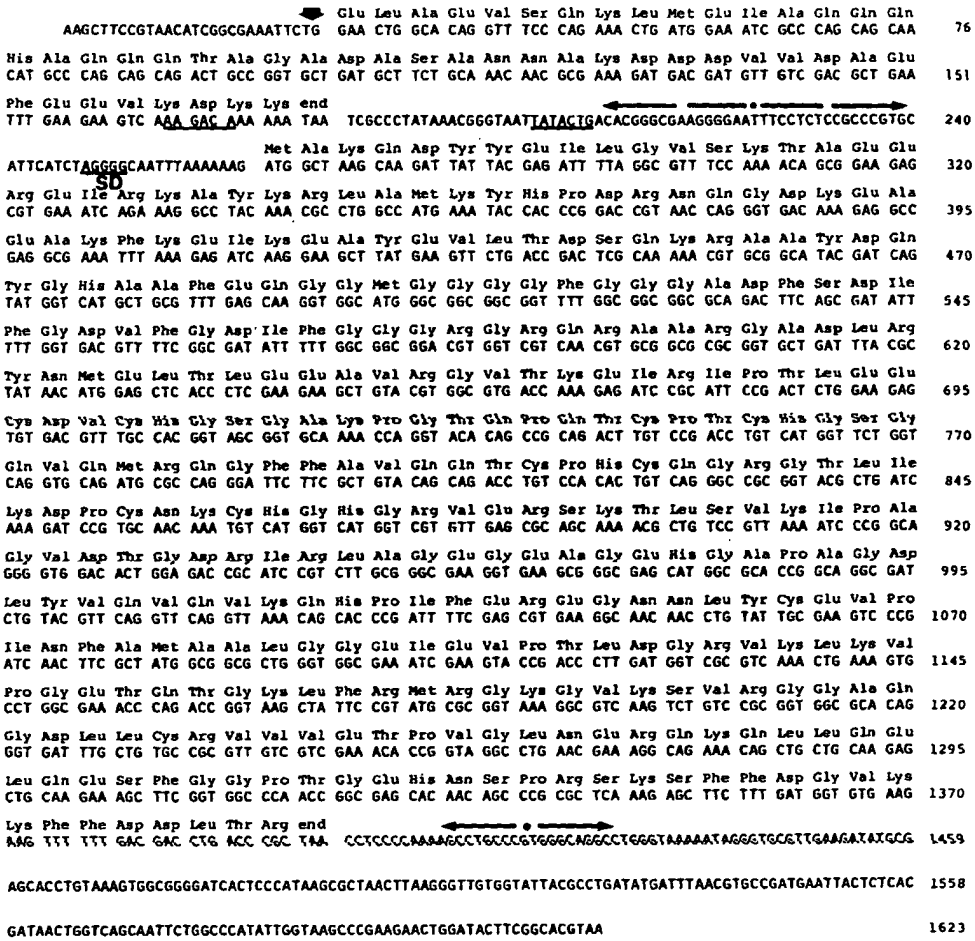

FIG. 2. **DNA sequence of the *dnaJ* gene and its flanking regions.** The position indicated by the *vertical arrow* shows the site of the Δ27 deletion introduced into the λ*dnaKdnaJ* phage (6). The inverted repeats in the DNA sequence are indicated by *horizontal arrows*. The putative promoter and ribosome binding sequences are *underlined*. The open reading frame is underscribed by corresponding amino acid sequences.

DISCUSSION

In this study, we described the nucleotide sequence of the *dnaJ* gene of *E. coli*, purification of the overproduced *dnaJ* protein, and determination of its NH_2-terminal amino acid sequence.

The initiation codon for *dnaJ* is located 88 bp downstream of the termination codon of the *dnaK* gene. Saito and Uchida (6) have reported that transcription of the *dnaJ* gene is initiated either from the promoter for the *dnaK* gene or from a weak promoter immediately upstream of the *dnaJ* gene. The most likely candidate for the second promoter sequence is TATACTG corresponding to the −10 region, located between nucleotides 223 and 229 (Fig. 2). However, assuming that this is the −10 region, sequences corresponding to the −35 region are not obvious. AAGACA, commonly accepted as the consensus sequence for the −35 region (22), appears at positions 189–194. No typical ribosome binding sequence is found at the expected position upstream from the given initiation codon of *dnaJ*. The sequence AGGGG 14–18 bases upstream from the initiation codon might function as the site of ribosome binding. Another feature we have noticed about the region upstream from the *dnaJ* gene was the existence of a strong palindromic structure, as indicated in Fig. 2.

The structure might function as an attenuation, resulting in reduced synthesis of the *dnaJ* protein. In constructing an operon fusion, pPL-*dnaJ*-23, to overproduce the *dnaJ* protein, sequences preceding the *dnaJ* gene were deleted to varying extents with endonuclease *Bal*31, and the clone producing the maximum amount of the *dnaJ* protein was selected. During the course of the selection, we have noticed that different recombinant plasmids constructed in this manner synthesized the protein in different amounts. Detailed analysis of the structure of these recombinant plasmids might provide an insight into the role of the palindromic region mentioned above and the signal responsible for the reduced production of the *dnaJ* protein as compared to the abundant production of the *dnaK* protein in *E. coli*.

Another palindromic region 12 bases downstream from the termination codon of the *dnaJ* gene was noticed. It may be intriguing to speculate that the region might be a termination signal for mRNA synthesis. However, there is no T cluster immediately downstream from the palindrome as is seen in typical ρ-independent termination signals.

Edman degradation of the purified *dnaJ* protein indicated that the terminal methionine residue is processed *in vivo*. Therefore, the processed *dnaJ* protein is composed of 375 amino acid residues, and its molecular weight is calculated to be 40,975. The amino acid composition shows that the protein

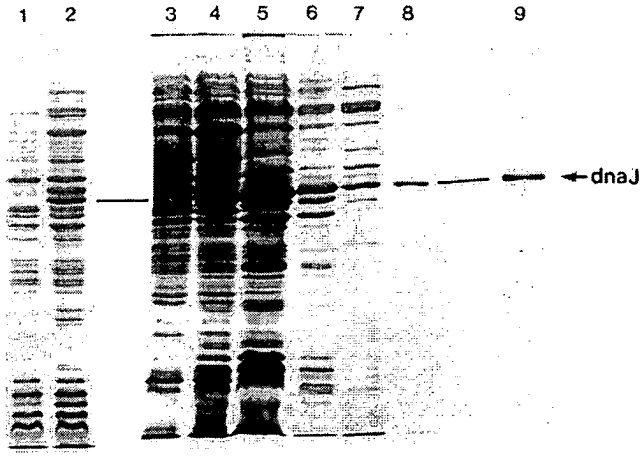

FIG. 3. **Purification of the *dnaJ* protein overproduced in N4830 cells transformed by pPL-*dnaJ*-23.** Cells were induced and extracts were fractionated as described in "Experimental Procedures." Aliquots of fractions at each step were analyzed by Na-DodSO$_4$-polyacrylamide gel electrophoresis and stained with Coomassie Brilliant Blue. *Lanes 1* and *2* are the proteins in uninduced and induced cells, respectively. *Lanes 3–5* correspond to the proteins in the lysozyme supernatant, fluffy precipitate, and tight precipitate of the centrifuged samples (see text), respectively. *Lane 6* is the 1 M sodium chloride extract of the *lane 5* sample. *Lanes 7* and *8* are the supernatant and precipitate fractions, respectively, of the sodium chloride extract after dialysis against a buffer containing a low concentration of the salt. *Lane 9* is the protein eluted from the acrylamide gel band.

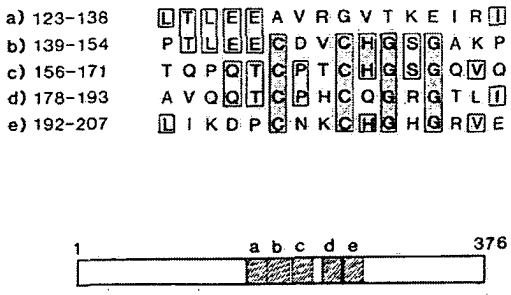

FIG. 4. **Homologous amino acid sequences tandemly duplicated within the *dnaJ* protein, as predicted from the DNA sequence of the gene (Fig. 2).**

is very rich in glycine (53 glycine residues) and basic amino acids (27 arginine, 27 lysine, and 10 histidine residues). A computer survey for internal homology of the amino acid sequence of the protein revealed that a segment composed of 16 amino acid is duplicated tandemly four or five times (Fig. 4). A search of the protein sequence data base (NBRF/PIR) disclosed that similar tandem duplication of the sequence Cys··Cys·Gly·Gly also occurs within the regulatory protein Q of the λ phage between amino acid residues 128 and 161 of the protein. However, the biological meaning of the occurrence of the sequence is not clear at present.

Acknowledgments—We are grateful to Dr. Motoo Tomita for determining the NH_2-terminal amino acid sequence and Drs. Koichi Titani and Koji Takio for computer analysis of the DNA sequence data. We are also indebted to Fumiko Usuda for assistance in protein purification and to Dr. Pat Lund for language assistance in preparation of the manuscript.

REFERENCES

1. Georgopoulos, C. P., and Herskowitz, I. (1971) in *The Bacteriophage Lambda* (Hershey, A. D., ed) pp. 553–564, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2. Furth, M. E., and Wickner, S. H. (1983) in *Lambda II* (Hendrix, R., Roberts, J., Stahl, F., and Weisberg, R., eds) pp. 145–173, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
3. Saito, H., and Uchida, H. (1977) *J. Mol. Biol.* **113,** 1–25
4. Sunshine, M., Feiss, M., Stuart, J., and Yochem, J. (1977) *Mol. Gen. Genet.* **151,** 27–34
5. Yochem, J., Uchida, H., Sunshine, M., Saito, H., Georgopoulos, C. P., and Feiss, M. (1978) *Mol. Gen. Genet.* **164,** 9–14
6. Saito, H., and Uchida, H. (1978) *Mol. Gen. Genet.* **164,** 1–8
7. Itikawa, H., and Ryu, J.-I. (1979) *J. Bacteriol.* **138,** 339–344
8. Georgopoulos, C. P., Tilly, K., Drahos, D., and Hendrix, R. (1982) *J. Bacteriol.* **149,** 1175–1177
9. Bardwell, J. C. A., and Craig, E. A. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 848–852
10. Zylicz, M., LeBowitz, J. H., McMacken, R., and Georgopoulos, C. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6431–6435
11. Tamura, F., Kanazawa, H., Tsuchiya, T., and Futai, M. (1981) *FEBS Lett.* **127,** 48–52
12. Reed, R. R. (1981) *Cell* **25,** 713–719
13. Georgopoulos, C. P., Lindquist-Heil, A., Yochem, J., and Feiss, M. (1980) *Mol. Gen. Genet.* **178,** 583–588
14. Tamura, F., Nishimura, S., and Ohki, M. (1984) *EMBO J.* **3,** 1103–1107
15. Fuller, R. S., Kaguni, J. M., and Kornberg, A. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 7370–7374
16. Hunkapiller, M. W., Lujan, E., Ostrander, F., and Hood, L. E. (1983) *Methods Enzymol.* **91,** 227–236
17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
18. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78
19. Ames, G. F.-L. (1974) *J. Biol. Chem.* **249,** 634–644
20. Georgopoulos, C. P. (1977) *Mol. Gen. Genet.* **151,** 35–39
21. Daniels, D., Schroeder, J., Szybalski, W., Sanger, F., Coulson, A., Hong, G., Hill, D., Peterson, G., and Blattner, F. (1983) in *Lambda II* (Hendrix, R., Roberts, J., Stahl, F., and Weisberg, R., eds) pp. 519–676, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
22. Rosenberg, M., and Court, D. (1979) *Annu. Rev. Genet.* **13,** 319–353